

University of Groningen

Overcoming the Problems of Inconsistent International Migration data

de Beer, Joop; Raymer, James; van der Erf, Rob; van Wissen, Leo

Published in:

European journal of population-Revue europeenne de demographie

DOI:

[10.1007/s10680-010-9220-z](https://doi.org/10.1007/s10680-010-9220-z)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

de Beer, J., Raymer, J., van der Erf, R., & van Wissen, L. (2010). Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe. *European journal of population-Revue europeenne de demographie*, 26(4), 459-481. <https://doi.org/10.1007/s10680-010-9220-z>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Overcoming the Problems of Inconsistent International Migration data: A New Method Applied to Flows in Europe

Surmonter les problèmes d'incohérences des données sur les migrations internationales: une nouvelle méthode appliquée aux flux en Europe

Joop de Beer · James Raymer · Rob van der Erf ·
Leo van Wissen

Received: 23 March 2010 / Accepted: 7 July 2010 / Published online: 28 October 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Due to differences in definitions and measurement methods, cross-country comparisons of international migration patterns are difficult and confusing. Emigration numbers reported by sending countries tend to differ from the corresponding immigration numbers reported by receiving countries. In this paper, a methodology is presented to achieve harmonised estimates of migration flows benchmarked to a specific definition of duration. This methodology accounts for both differences in definitions and the effects of measurement error due to, for example, under reporting and sampling fluctuations. More specifically, the differences between the two sets of reported data are overcome by estimating a set of adjustment factors for each country's immigration and emigration data. The adjusted data take into account any special cases where the origin–destination patterns do not match the overall patterns. The new method for harmonising migration flows that we present is based on earlier efforts by Poulain (European Journal of Population, 9(4): 353–381 1993, Working Paper 12, joint ECE-Eurostat Work Session on Migration Statistics, Geneva, Switzerland 1999) and is illustrated for movements between 19 European countries from 2002 to 2007. The results

J. de Beer (✉) · R. van der Erf
Netherlands Interdisciplinary Demographic Institute (NIDI), PO Box 11650,
2502 AR The Hague, The Netherlands
e-mail: beer@nidi.nl

J. Raymer
Southampton Statistical Sciences Research Institute (S3RI), University of Southampton,
Highfield, Southampton SO17 1BJ, UK

L. van Wissen
Population Research Centre, Faculty of Spatial Sciences, University of Groningen,
PO Box 800, 9700 AV Groningen, The Netherlands

represent a reliable and consistent set of international migration flows that can be used for understanding recent changes in migration patterns, as inputs into population projections and for developing evidence-based migration policies.

Keywords Immigration · Emigration · International migration statistics · Harmonisation · Europe

Résumé Les comparaisons entre pays des caractéristiques des migrations internationales s'avèrent difficiles et confuses du fait des différences de définitions et de méthodes de mesure. Les chiffres de l'émigration fournis par les pays d'origine des migrants diffèrent souvent des chiffres recueillis par les pays d'accueil. Cet article propose une méthode destinée à rendre cohérentes les estimations des flux migratoires en se référant à une définition spécifique de la durée de séjour. Cette méthodologie tient compte à la fois des différences de définitions et des conséquences des erreurs de mesure dues, par exemple, au sous-enregistrement et aux fluctuations de l'échantillonnage. Plus spécifiquement, les différences entre les deux séries de données rapportées sont corrigées en estimant un ensemble de facteurs d'ajustement des données pour chaque pays d'immigration et d'émigration. Les données ajustées tiennent compte de tous les cas particuliers dans lesquels les schémas origine–destination ne correspondent pas aux schémas généraux. Cette nouvelle méthode d'harmonisation des données des flux migratoires, basée sur les travaux antérieurs de Poulain (1993, 1999), est appliquée aux mouvements migratoires entre 19 pays européens de 2002 à 2007. Les résultats présentent des séries fiables et cohérentes de flux migratoires internationaux qui peuvent être utilisées pour comprendre les changements récents dans les schémas migratoires, pour réaliser des projections de population, et pour développer des politiques migratoires à partir d'informations plus sûres.

Mots-clés Immigration · Émigration · Statistiques de migrations internationales · Harmonisation · Europe

1 Introduction

Our understanding of the mechanisms and patterns of international migration over time are impeded both by the lack of data and by inconsistencies in the measurement and collection of the data that are available. In fact, it is well known that the patterns of migration vary significantly depending on which country is reporting the data (Kupiszewska and Nowok 2008; Nowok et al. 2006; Zlotnik 1987). Considering that international migration is the main factor contributing to population growth in Europe, this is very unfortunate. In response to the problem of inconsistent migration data, we have developed a methodology for harmonising the data available to us from countries in Europe. More specifically, we make use of doubly counted information obtained from migrant sending and migrant receiving countries to estimate adjustment factors necessary for producing a consistent set of migration flows. These estimated flows are benchmarked to a particular definition.

Harmonisation of migration data is required for the development of policies on immigration (Kraier et al. 2006). Differences in both the concepts and techniques used to measure migration make any international comparison of migration difficult. There has been a lot of work on data issues and migration definitions, for example, see Champion (1994), Kelly (1987), Kraly and Gnanasekaran (1987), Poulain (1993), Poulain et al. (2006), Raymer and Willekens (2008), United Nations (2002) and Willekens (1994, 1999). Several international institutes such as the *International Labour Organization*, the *Organisation for Economic Co-operation and Development*, the *United Nations* and the *European Commission* have all invested heavily in the harmonisation of international migration data, but without much success or progression (Bilsborrow et al. 1997; Herm 2006a; Fassmann 2009). In fact, the situation today in terms of migration definitions and measurement is not much better than it was, say, 20 years ago.

Recently, some renewed efforts have been made to improve the migration data situation in Europe. In 2007, the European Parliament adopted a new regulation on migration statistics. This regulation provides clear definitions of immigration and emigration (Official Journal of the European Union 2007), and lists the migration data that must be supplied to *Eurostat*, the statistical office of the European Union (EU), by Member States. However, this regulation leaves the Member States free to decide how they will provide these data, including the use of estimation methods (Fassmann 2009). The methodology presented in this paper should help national statistical offices to improve and harmonise the data they currently provide to international organisations, such as Eurostat.

The migration definition set out in the 2007 Regulation corresponds to the definition recommended by the United Nations (1998), where an international migrant is defined as ‘a person who moves to a country other than that of his or her usual residence for a period of at least a year’. One problem affecting the implementation of this definition is that some countries are unable to identify their nationals who have left (Fassmann 2009). Furthermore, many European countries exclude the immigration of nationals from the published statistics, as they are not considered to be ‘migrants’. Another important obstacle has to do with the recommended duration of residence in the country of destination. It may take up to 2 years to identify all persons who have stayed at least 1 year, as they may arrive anytime during the annual time period of interest. This means that the publication of migration statistics based on the actual duration of stay may be delayed for some time. To provide statistics to the user community in a quicker fashion, many countries simply count those migrants who have stayed for at least 3 months, which leads to higher numbers than if the 1-year criterion was applied. Other countries use the intended duration of stay as the criterion (Fassmann 2009).

Many European countries do not have reliable statistics on emigration. This is mainly caused by the fact that migrants have little incentive to report their move to the administration of the country they have emigrated from. Moreover, it is difficult to count persons leaving the country because they are no longer present in the country collecting the data. In this situation, comparisons of sending country data with receiving country data provide important information on the degree of underestimation found in reported emigration flows (UNECE 2009). In fact, the analysis of the so-

called ‘double-entry matrix’ of migration flows produced by UNECE since the early 1970s, and more recently by Eurostat, has been found to be very useful and informative. Kelly (1987) and Poulain (1999), for example, have used the information contained in this matrix to assess the degree of harmonisation amongst reported data. In doing so, the possibility that very narrow or loose definitions of migration may be used for reported immigration statistics must be taken into account, which results in lower or higher levels of migration flows, respectively, in relation to, say, the United Nations’ recommended 1-year definition (UNECE 2009).

The aim of this paper is to illustrate how reliable estimates of harmonised migration statistics may be obtained from a set of origin–destination flows, where two reported flows are available for each particular flow, i.e., from receiving and sending countries. The new method that we present is based on earlier efforts by Poulain (1993, 1999), and is applied to reported flows between 19 European countries from 2002 to 2007. Note, however, that this paper does not consider flows outside the 19 country system, or those that are missing. Raymer (2008) describes a method for estimating missing migration flow data.

2 Comparability of International Migration Data

The reliability of migration statistics can be measured by how well they correspond to a particular country’s definition or concept of migration. However, as definitions differ across countries, reliability does not guarantee comparability. Moreover, under-registration, under-coverage and accuracy of the collection system also affect the measurement of migration (Bilsborrow et al. 1997; Nowok et al. 2006). First, there may be under-registration of migrants. This may be the case if the data depend on declarations by the migrants themselves. The willingness to report changes in places of residence varies both between countries and between groups of migrants. In general, migrants have more incentive to report their arrival than their departure, as there are usually direct benefits in doing so (e.g., access to social services). Therefore, immigration statistics are generally considered more reliable than emigration statistics (Thierry et al. 2005; UNECE 2009). Second, there may be under-coverage. This measurement category refers to the non-inclusion of particular migrant groups. Here, the differences are most often caused by the absence or inclusion of nationals, students, asylum seekers or irregular (illegal) migrants in the data. In general, asylum seekers are included only when they have been granted refugee status and received a temporary or permanent residence permit. However, in some instances, they are registered at an earlier stage of the asylum process. In other instances, even recognised refugees are not included. Irregular migrants are generally not included in migration statistics, as they are especially difficult to measure (for obvious reasons). In fact, Spain is the only EU country that includes irregular migrants in the official statistics. Finally, data based on sample surveys may be unreliable due to sampling errors. Furthermore, unless the sample size is very large, the data are likely to show irregularities in the patterns across ages or in the distribution of origins or destinations over time, as flows of migrants represent a relatively small proportion of the overall population being surveyed.

The main sources of the differences in the definitions used by EU countries to measure migration are the concepts of place of residence and duration of stay (Zlotnik 1987; Bilsborrow et al. 1997; Kupiszewska and Nowok 2008). The *de jure* (legal) approach to residence implies that in order to become a resident, a migrant must comply with certain regulations, which tend to differ between nationals and foreigners, and among foreigners, between EU- and non-EU-nationals. For example, it is not uncommon for emigrants to be registered in their country of citizenship (origin) even after several years of living abroad (Thierry et al. 2005). Thus, having a place of residence does not necessarily imply a presence in that country. The *de facto* (actual) approach is connected with physical presence in a country, usually for a specified minimum period of time. To prevent the delay caused by measuring actual duration of stay, most European countries use the intended duration of stay instead (Nowok et al. 2006). Alternatively, the intended duration of stay may be used to provide provisional statistics, which are updated at a later point with the actual duration of stay statistics. Another group of countries measure ‘permanent’ change of residence only (e.g. Poland and Slovakia), which is very restrictive and tends to produce flow levels that are much lower relative to other definitions. The duration of stay criterion used by the majority of EU countries is between 3 months and 1 year. Only three countries (Cyprus, Sweden and UK) apply strictly the 1-year criterion for immigration, as well as for emigration and for both nationals and non-nationals (Thierry et al. 2005). In fact, some countries do not take duration of stay into account at all. Germany is such an example, where everybody taking up a residence is counted as a migrant.

Because of differences in definition, coverage, registration and accuracy of the collection mechanism, the origin–destination matrix of migration flows between European countries based on immigration data reported by the countries of origin tends to differ from the matrix reported by the countries of destination. With respect to definitions, the differences are expected to be systematic over time. For example, the German definition is wider than the Dutch definition which, in turn, is wider than that of Sweden. In fact, Germany reports higher figures than the Netherlands, and the figures of the Netherlands are higher than those reported by Sweden (Kupiszewska and Nowok 2008). A comparison of the size of these reported flows provides information on the effects of differences in definition on the size of migration flows (Bilsborrow et al. 1997; UNECE 2009). However, as mentioned above, not all differences can be explained by differences in definition. In some cases, countries report relatively large percentages of unknown countries of origin or destination. Furthermore, sudden jumps in observations may be caused by changes in definitions or by changes in the registration method.

Data on immigration and emigration flows by country of origin and destination are usually presented in an origin–destination matrix with off diagonal entries containing the number of people moving from any origin i to any destination j in a given calendar year. For this study, we have collected migration data for the 19 countries set out in Table 1. As each flow can be reported by both sending and receiving countries, two migration tables may be produced. Such data are set out in Tables 2 and 3. Here, the average 2002–2007 values of migration between the 19 European countries set out in Table 1 are presented. Table 2 contains flows reported

Table 1 List of European countries reporting both immigration flows by country of origin and emigration flows by country of destination, 2002–2007

Country	Abbreviation
Austria	AT
Cyprus	CY
Czech Republic	CZ
Germany	DE
Denmark	DK
Spain	ES
Finland	FI
Iceland	IS
Italy	IT
Lithuania	LT
Luxembourg	LU
Latvia	LV
Netherlands	NL
Norway	NO
Poland	PL
Sweden	SE
Slovenia	SI
Slovakia	SK
United Kingdom	UK

by the countries of destination and Table 3 contains the flows reported by the countries of origin. Clearly, there are large differences between the two sets of reported numbers (see, e.g., Spain to the United Kingdom or Poland to Germany).

3 Method

The differences between reported immigration and emigration numbers are useful for improving and harmonising the migration data. If reported emigration numbers for a given country turn out to be systematically lower than the corresponding immigration numbers reported by the countries of destination, this suggests that the reported emigration numbers are too low. Adjusting these numbers in an upward direction moves them closer to the actual numbers. The same applies to reported immigration numbers. For each country we can estimate one adjustment factor for immigration and one for emigration in such a way that the adjusted immigration and emigration numbers are closer to each other than the reported numbers. To prevent arbitrary judgments biasing the results, we believe the estimation of adjustment factors for immigration and emigration flows should be estimated simultaneously. Moreover, it should be noted that immigration is not necessarily recorded more accurately than emigration. In some situations, sending country data may be considered better (Nowok et al. 2006).

Poulain (1993, 1999) was the first to develop a method to adjust reported immigration and emigration numbers for the purpose of obtaining a consistent set of

Table 2 Reported migration by country of destination, averages 2002–2007

From	To									
	AT	CY	CZ	DE	DK	ES	FI	IS	IT	LT
AT		41	310	14,257	303	774	109	33	774	17
CY	22		13	276	23	25	23	1	30	3
CZ	1,316	118		9,218	262	833	56	42	672	24
DE	15,447	332	1,362		4,001	15,982	921	255	12,809	490
DK	203	25	46	2,687		964	365	1,413	265	85
ES	700	45	71	14,703	1,758		644	68	2,044	252
FI	270	21	38	2,173	414	844		45	235	43
IS	31	0	4	236	1,665	131	50		35	10
IT	1,608	49	254	22,196	986	9,320	250	74		82
LT	179	35	47	4,496	1,034	2,274	73	272	378	
LU	67	3	2	2,282	162	123	50	27	213	5
LV	83	104	13	2,155	457	300	87	93	183	175
NL	791	70	255	13,681	864	4,762	261	55	905	41
NO	98	14	24	1,378	3,148	1,696	845	364	167	87
PL	5,231	752	1,608	136,927	2,436	8,277	187	2,229	9,045	120
SE	489	88	67	3,348	3,313	1,826	3,502	492	379	91
SI	556	9	17	1,798	46	136	6	9	321	2
SK	3,192	432	14,064	11,148	149	788	22	45	690	4
UK	1,222	3,170	506	13,263	3,482	38,674	946	228	4,553	875
Total	31,504	5,306	18,702	256,221	24,502	87,725	8,397	5,741	33,695	2,407

From	To									
	LU	LV	NL	NO	PL	SE	SI	SK	UK	Total
AT	8	9	559	111	180	307	100	208	1,395	19,496
CY	0	2	51	15	7	61	2	2	2,533	3,087
CZ	4	15	511	116	45	164	6	979	4,109	18,489
DE	454	166	9,182	2,268	2,876	3,374	299	446	19,039	89,701
DK	11	46	475	2,943	34	5,264	3	21	1,874	16,721
ES	24	18	3,101	768	119	1,300	8	36	14,581	40,239
FI	3	43	379	799	6	3,204	1	6	684	9,208
IS	0	6	75	373	11	462	1	2	417	3,509
IT	67	33	1,811	246	309	599	79	109	5,829	43,900
LT	1	236	302	926	43	574	0	5	2,507	13,380
LU		2	161	18	5	90	5	1	682	3,897
LV	2		125	233	6	264	0	5	1,227	5,511
NL	27	20		711	163	979	12	41	6,799	30,436
NO	2	24	453		48	5,098	1	24	1,667	15,135
PL	19	45	5,744	4,602		3,718	3	276	36,759	217,977
SE	14	54	696	4,917	113		15	20	3,213	22,635
SI	1	1	90	14	2	42		16	0	3,064
SK	4	4	465	238	18	110	4		4,584	35,961
UK	39	190	5,820	1,624	1,126	3,114	22	116		78,969
Total	682	913	30,000	20,921	5,111	28,723	559	2,311	107,897	671,315

Source: Eurostat

Table 3 Reported migration by country of origin, averages 2002–2007

From	To									
	AT	CY	CZ	DE	DK	ES	FI	IS	IT	LT
AT		18	937	6,665	166	429	231	27	1,022	111
CY	6		21	57	6	19	12	0	39	9
CZ	186	13		560	24	35	28	2	112	10
DE	17,787	271	8,104		3,095	16,807	2,371	287	31,235	2,455
DK	228	24	179	2,612		1,669	368	1,347	716	655
ES	155	9	57	2 686	157		110	9	1 163	120
FI	97	23	42	758	400	671		53	203	21
IS	13	2	17	205	1,800	59	48		105	64
IT	588	6	67	10,206	149	1,508	136	17		11
LT	48	8	54	1,269	158	628	87	23	204	
LU	31	3	13	911	99	79	35	19	175	4
LV	18	8	6	302	45	18	46	5	51	138
NL	616	50	298	10,493	533	3,774	322	54	1,278	54
NO	69	17	43	709	3,093	789	855	412	146	108
PL	538	15	63	14,417	111	341	20	46	505	6
SE	298	73	104	1,634	3,159	1,348	3,403	413	463	48
SI	311	3	14	589	5	27	4	1	186	1
SK	177	1	629	255	4	16	1	0	42	0
UK	1,593	4,060	2,692	12,579	1,932	33,431	682	103	5,270	1,074
Total	22,758	4,600	13,339	66,905	14,933	61,649	8,758	2,818	42,914	4,887

From	To									
	LU	LV	NL	NO	PL	SE	SI	SK	UK	Total
AT	45	42	426	87	2,401	388	402	1,778	901	16,076
CY	2	18	10	2	111	13	0	32	371	724
CZ	3	7	81	16	583	24	8	9,539	219	11,449
DE	1,686	1,494	9,293	2,122	100,827	3,974	2,004	9,456	17,233	230,499
DK	138	316	602	2,947	833	5,253	31	95	3,889	21,898
ES	87	19	869	159	398	203	10	45	3 430	9 684
FI	71	27	233	777	63	3,216	4	10	1,175	7,842
IS	37	29	49	482	872	478	25	56	232	4,570
IT	218	8	531	121	417	199	151	40	3,508	17,879
LT	18	163	116	199	122	233	3	5	2,638	5,975
LU		4	97	12	23	73	5	11	166	1,760
LV	6		20	34	26	67	1	2	196	987
NL	191	33		731	1,020	900	45	138	7,953	28,482
NO	23	69	287		281	5,083	5	61	1,395	13,444
PL	23	3	557	127		303	2	10	5,219	22,306
SE	127	62	522	4,746	354		27	29	3,905	20,713
SI	24	0	30	5	5	38		6	70	1,319
SK	2	0	13	3	15	8	3		69	1,235
UK	362	324	5,943	1,993	6,507	2,666	0	1,053		82,264
Total	3,062	2,619	19,676	14,561	114,854	23,117	2,724	22,364	52,567	499,105

Source: Eurostat

migration flows. ‘Correction factors’ were estimated by minimising the sum of squares $\sum_{i,j} (\hat{\alpha}_j I_{ij} - \hat{\beta}_i E_{ij})^2$, where I_{ij} denotes migration from country i to country j reported by the receiving country j , E_{ij} denotes the same flow reported by the sending country i , α_j is the adjustment factor for all immigration to country j and β_i is the adjustment factor for all emigration from country i . Poulain and Dal (2008) refined this method by dividing the squared differences by the sum of the reported numbers, i.e.,

$$\sum_{i,j} (\hat{\alpha}_j I_{ij} - \hat{\beta}_i E_{ij})^2 / (I_{ij} + E_{ij}). \quad (1)$$

This refinement prevents flows from (or to) large countries from biasing the estimates.

Various constraints have been tried by Poulain and colleagues (Abel 2009). For instance, following the iterative approach to harmonising migration flows suggested by van der Erf and van der Gaag (2007), Poulain and Dal (2008) proposed that the estimates should be normalised to Swedish immigration data, as they are generally considered to be highly reliable and in agreement with the UN recommended measure, as well as with the new EU regulation (Herm 2006b). The parameters α_j and β_i may be estimated by solving a system of linear equations, which result from applying the method of Lagrange multipliers. Multiplying I_{ij} by $\hat{\alpha}_j$ and E_{ij} by $\hat{\beta}_i$ produces two sets of migration flow estimates from country i to country j . The final set of estimates are obtained by simply taking the average of the two, i.e., $\hat{n}_{ij} = (\hat{\alpha}_j I_{ij} + \hat{\beta}_i E_{ij})/2$, where \hat{n}_{ij} denotes the harmonised migration flows. Note, Poulain and Dal (2008) applied their correction method first to countries with relatively reliable data to prevent countries with less reliable data influencing the overall patterns. Here, the main concern is that the less reliable data have origin–destination patterns that are not consistent with the actual patterns. Thus, less reliable flows were adjusted in a hierarchical fashion, i.e., by using the harmonised reliable data as a basis.

There are several limitations in the model described above. First, the reported numbers included in the denominator of Eq. 1 are known to be incorrect (Abel 2009). Second, the row and column totals of the two estimated matrices are not equal. As a result, the row and column totals of the average harmonised migration matrix do not correspond to the row and column totals estimated using the adjustment factors. Finally, the method can only be applied to a limited set of countries with reasonably reliable data. This implies that the estimates of the adjustment factors depend on the selection of countries, which may not reflect the broader patterns of interest. For these reasons, we have revised Poulain’s method in two important ways. First, the row-sums and column-sums of the two estimated matrices are set to be equal. Second, we introduce additional constraints on individual cells in the migration matrices, so that more countries (with less reliable data) may be included.

The adjustment factors for our method can be estimated by solving a system of linear equations and imposing a constraint. If we have a $N \times N$ receiving country and an equivalent $N \times N$ sending country matrix, the adjustment factors for

receiving country, α_j , and the adjustment factors for sending country data, β_i , can be estimated by

$$\sum_j \hat{\alpha}_j I_{ij} = \hat{\beta}_i \sum_j E_{ij} \quad \text{for } i = 1, \dots, N; \quad i \neq j \quad (2)$$

$$\hat{\alpha}_j \sum_i I_{ij} = \sum_i \hat{\beta}_i E_{ij} \quad \text{for } j = 1, \dots, N; \quad i \neq j \quad (3)$$

Equation 2 states that for each country the emigration total estimated on the basis of the adjusted matrix of flows reported by receiving countries equals the emigration total estimated on the basis of the adjusted matrix of flows reported by sending countries. Equation 3 does the same for immigration totals.

Equations 2 and 3 can be written as a homogeneous system of $2N$ linear equations with $2N$ unknowns, i.e.,

$$\begin{aligned} \hat{\alpha}_2 I_{12} + \hat{\alpha}_3 I_{13} + \dots + \hat{\alpha}_N I_{1N} - \hat{\beta}_1 \sum_j E_{1j} &= 0 \\ &\vdots \\ \hat{\alpha}_1 I_{N1} + \hat{\alpha}_2 I_{N2} + \dots + \hat{\alpha}_{N-1} I_{NN-1} - \hat{\beta}_N \sum_j E_{Nj} &= 0 \\ \hat{\alpha}_1 \sum_i I_{i1} - \hat{\beta}_2 E_{21} - \hat{\beta}_3 E_{31} - \dots - \hat{\beta}_N E_{N1} &= 0 \\ &\vdots \\ \hat{\alpha}_N \sum_i I_{iN} - \hat{\beta}_1 E_{N1} - \hat{\beta}_2 E_{N2} - \dots - \hat{\beta}_{N-1} E_{NN-1} &= 0 \end{aligned} \quad (4)$$

This system has an infinite number of solutions for α_j and β_i . For each set of values of α_j and β_i that solve this system, $k\hat{\alpha}_j$ and $k\hat{\beta}_i$ are solutions as well. In order to find a unique solution one restriction needs to be imposed. In accordance with Poulain and Dal (2008), we assume that the adjustment factor for Swedish immigration is equal to one, since Sweden uses a definition of migration that is consistent with the new EU regulation and the quality of Swedish immigration data is considered to be adequate. This also means that the resulting estimates are harmonised in line with the new European regulation.

The basic assumption underlying our estimation procedure (as described above) is that the distributions of reported immigration by country of origin and reported emigration by country of destination correspond to the distribution of actual migration flows under the harmonised definition. This implies that the reported emigration of country A is $x\%$ higher or lower than the actual number (based on the standard definition) for all countries of destination. The same assumption applies to receiving country numbers. However, as we find in the next section, the estimated receiving country flows by country of origin and the estimated sending country flows by country of destination are not always consistent with each other. In a number of cases, specific origin–destination flows have to be considered separately. For that reason, we introduce additional constraints, corresponding to particular origin–destination flows that differ from the remaining flows.

Let us assume that the estimated receiving country migration flow from country p to q , $\hat{\alpha}_q I_{pq}$, differs substantially from the estimated sending country flow, $\hat{\beta}_p E_{pq}$. To make them consistent, we can multiply $\hat{\alpha}_q I_{pq}$ by $\hat{\gamma}_{pq}$ or $\hat{\beta}_p E_{pq}$ by $\hat{\delta}_{pq}$ so that both estimates of migration are equal. The question whether we should adjust the estimate based on the reported receiving country or the estimate based on the reported sending country depends on our knowledge of the data.

Given the estimated values of $\hat{\alpha}_q$ and $\hat{\beta}_p$ we can calculate the value of $\hat{\gamma}_{pq}$ easily from $\hat{\gamma}_{pq} = \hat{\beta}_p E_{pq} / \hat{\alpha}_q I_{pq}$ or the value of $\hat{\delta}_{pq}$ from $\hat{\delta}_{pq} = \hat{\alpha}_q I_{pq} / \hat{\beta}_p E_{pq}$. However, introducing $\hat{\gamma}_{pq}$ or $\hat{\delta}_{pq}$ changes the estimates of $\hat{\alpha}_q$ or $\hat{\beta}_p$. This also means that the row and column totals of both estimated migration matrices no longer tally. Therefore, we adjust the system of linear Eqs. 2 and 3 by adding constraints on individual cells of the matrices. If we assume that the emigration number reported by country p needs to be adjusted, Eqs. 2 and 3 can be rewritten as

$$\sum_j \hat{\alpha}_j I_{ij} = \hat{\beta}_i \sum_j E_{ij} (1 + \hat{\delta}_{pq}^* D_{ij}) \quad \text{for } i = 1, \dots, N; \quad i \neq j \quad (5)$$

$$\hat{\alpha}_j \sum_i I_{ij} = \sum_i \hat{\beta}_i E_{ij} (1 + \hat{\delta}_{pq}^* D_{ij}) \quad \text{for } j = 1, \dots, N; \quad i \neq j \quad (6)$$

where $D_{ij} = 1$ if $i = p$ and $j = q$, $D_{ij} = 0$ otherwise, and $\hat{\delta}_{pq}^* = \hat{\delta}_{pq} - 1$.

The equations including I_{pq} and E_{pq} in the system of Eq. 4 can be rewritten as follows:

$$\hat{\alpha}_1 I_{p1} + \dots + \hat{\alpha}_q I_{pq} + \dots + \hat{\alpha}_N I_{pN} - \hat{\beta}_p E_{p1} - \dots - \hat{\delta}_{pq} \hat{\beta}_p E_{pq} - \dots - \hat{\beta}_p E_{pN} = 0 \quad (7)$$

$$\hat{\alpha}_q \sum_i I_{iq} - \hat{\beta}_1 E_{1q} - \dots - \hat{\delta}_{pq} \hat{\beta}_p E_{pq} - \dots - \hat{\beta}_N E_{Nq} = 0$$

In contrast with Eq. 4, these are non-linear equations, because they include the term $\hat{\delta}_{pq} \hat{\beta}_p E_{pq}$. The values of the coefficients can be estimated by an iterative procedure. The model can be extended in a straightforward way to include additional constraints. However, for any particular country, the number of constraints should not be too high, as this reduces the available information to estimate α and β .

4 Data

The sending and receiving country migration data have been provided by the national statistical institutes of the EU Member States in response to annual rounds of data collection conducted jointly by five international organisations and coordinated by Eurostat (Kupiszewska and Nowok 2008). As concerns Europe, Eurostat processes and disseminates data received from 37 countries on their website (epp.eurostat.ec.europa.eu). Data sources used by EU member states to produce migration statistics are very diverse (Kupiszewska and Nowok 2008;

Nowok et al. 2006). The major types of sources are population registration systems, statistical forms, other administrative registers related to foreigners (such as alien registers, residence permit registers and registers of asylum seekers), sample surveys and censuses. Thirteen EU countries use a population register as the source of migration statistics. Alien registers and residence permit registers are used in seven countries, sometimes in addition to population registers. These registers only provide information on the migration of non-nationals. Cyprus and the UK rely on passenger surveys conducted at the borders, while Portugal and Ireland rely on household surveys. Greece, France and Portugal do not have any data on migration by nationals. Some countries derive their emigration statistics from data on residence permits by assuming a migrant has left the country when a residence permit has expired. Moreover, they often assume that the country of next residence is the country of their citizenship. The result, we believe, is an overestimation of actual emigration to those particular countries. Finally, several countries include in their so-called ‘administrative corrections’ emigration that has not been declared, which cannot be disaggregated by country of next residence.

Data on immigration by country of previous residence or emigration by country of next residence are not always available or complete (Nowok et al. 2006). Thus, the sending country and receiving country matrices, when combined into a double-entry matrix may be incomplete. For some countries, a large share of emigrants have an unknown country of destination: around 75% in Slovenia, 40% in Luxembourg, 35% in Austria, 31% in the Netherlands and 39% in Spain, for example. Fortunately, the estimation of adjustment factors takes this into account.

In the next section, we present our harmonised estimates of migration between 19 European countries that provide data on both immigration by country of origin and emigration by country of destination for the calendar years 2002–2007. The reported data contains both nationals and non-nationals. Table 1 provides a list of the countries. Although there are some data for Ireland, Portugal and Romania, these have not been used because they cover only a part of the migration flows (e.g. only foreigners or nationals). For Iceland, Italy and Luxembourg, data for 1 or more years in the period 2002–2007 are missing. For these countries, the adjustment factors are estimated for averages over the available years.

5 Results

The results presented in this section are obtained by applying the estimation method described in Sect. 3. Table 2 shows the average values of migration between 19 European countries reported by receiving countries for the years 2002–2007 and Table 3 shows the corresponding numbers reported by the sending countries. The countries listed in the row headings refer to origins and those listed in the column headings refer to destinations. A comparison of Tables 2 and 3 reveals large differences between numbers reported by sending and receiving countries. According to the numbers reported by receiving countries, 671,315 migrants per year moved between these 19 countries, whereas the numbers reported by sending countries total 499,105. For 11 countries, the reported receiving country

immigration totals are higher than the corresponding sending country totals. For example, Germany reported that 256,221 immigrants arrived from the 18 countries in this study, whereas these countries reported that only 66,905 emigrants moved to Germany. Poland reported that 22,306 persons emigrated to the other 18 countries which, for their part, reported receiving 217,977 immigrants from Poland, suggesting that Polish emigration data are around 10 times too low. For 15 of the 19 countries, the emigration total reported by the sending country is lower than the corresponding totals reported by receiving countries. Keep in mind that receiving country data should not always be considered better than sending country data. Consider, for example, the flows from Poland to Germany in Tables 2 and 3. Here, Germany received an average of 136,927 migrants from Poland, whereas Poland reported that they only sent an average of 14,417. This difference could be explained by the duration criteria used by these countries, with Germany having a very loose definition (instant) and Poland having a very restrictive definition (permanent). So, in comparison with the harmonised definition of a 1 year period, Germany's reported number is considered too high and Poland's too low.

The estimated adjustment factors are set out in Table 4. We indicated above that in order to estimate the adjustment factors a restriction was introduced, i.e., the adjustment factor for Swedish immigration is set equal to one. For 16 of the 19 countries, the E_{ij} adjustment factor exceeds one, indicating that sending country numbers tend to be underestimated. However, Table 4 also shows that I_{ij} numbers seem to be underestimated in the majority of countries as well. This may seem contradictory since for 11 of the 19 countries the reported immigration totals exceed

Table 4 Estimates of adjustment factors for immigration and emigration, 2002–2007

	Immigration	Emigration
Austria	1.06	1.74
Cyprus	1.06	5.29
Czech Republic	2.14	3.33
Germany	1.03	0.69
Denmark	0.74	0.80
Spain	0.82	4.90
Finland	1.26	1.22
Iceland	0.57	0.74
Italy	1.42	2.92
Lithuania	2.33	2.45
Luxembourg	5.65	2.43
Latvia	2.92	6.22
Netherlands	0.97	1.25
Norway	0.84	1.19
Poland	17.85	10.64
Sweden	1.00	1.21
Slovenia	5.18	2.71
Slovakia	18.90	43.69
United Kingdom	1.21	1.18

the corresponding emigration numbers reported by the sending countries. This is because the reported receiving country numbers should be compared with the adjusted sending country numbers rather than the reported numbers. For example, the immigration total reported by the UK (107,897) exceeds the reported emigration from sending countries to the UK (52,567). The reported emigration to the UK includes 5,219 emigrants from Poland to the UK. However, since the reported emigration from Poland is too low (the adjustment factor equals 18.31, see Table 4) the reported emigration from Poland to the UK is adjusted from 5,219 to 55,506. Moreover, the adjustment factor for Spanish emigration data equals 4.32, so the reported emigration from Spain to UK is adjusted from 3,430 to 16,792. For several other countries, emigration to the UK is adjusted upwards as well. As a consequence, the adjusted emigration numbers to the UK exceed the total of immigration reported by the UK and thus the reported immigration is adjusted upwards as well. Note that the adjustment factors for immigration for most countries are closer to one than the adjustment factors for emigration, which indicates that the reported immigration numbers are more accurate than the emigration numbers.

Multiplying the reported numbers in Table 2 by the adjustment factors for receiving country data and the reported numbers in Table 3 by the adjustment factors for sending country data results in two tables for which the row and column totals are equal (not presented here for space reasons). The differences between the cells in these two matrices are considerably smaller than those in Tables 2 and 3. In fact, the root mean squared error (RMSE) is reduced from 8,966 to 2,131. In other words, the differences between the two reported migration flow tables are reduced by 77%. However, we still found some substantial differences in the two estimated migration flow tables. For example, the migration from Poland to Germany estimated on the basis of German immigration data equals 141,035, whereas the estimate based on Polish emigration data is equal to 153,399. These differences reflect the fact that the distribution of reported Polish emigration by country of destination is not consistent with the share of immigration from Poland in the total reported immigration numbers of other countries. As a result, the estimate of the migration flow from Poland to Germany based on Polish data exceeds that based on German data, whereas for most other countries, the adjusted Polish emigration numbers are lower than the corresponding adjusted immigration numbers. This means that one substantial inconsistency in the estimates is likely to influence the estimates of other migration flows. To prevent such inconsistencies from affecting the overall estimates, we have added constraints to individual cells (flows) in the model.

The introduction of constraints to individual cells in the matrix allows us to consider special cases, such as the Poland to Germany flow described above. In total, we found six migration flows where the estimates differed by more than 10,000. Specifically, these flows were Poland to Germany, Poland to UK, Germany to Poland, Germany to UK, Czech Republic to Slovakia and UK to Poland. After identifying the flows with large differences, we then had to decide whether the constraint should be applied to the numbers of the receiving country or of the sending country. Since we believe that reported emigration numbers are generally considered to be less reliable than reported immigration numbers, we apply the

constraints to the sending country data, except for the Germany to Poland and UK to Poland flows (i.e., Poland's immigration data is considered to be of lower quality than the corresponding emigration data reported by both Germany and the UK).

The adjustment factors taking into account the six constraints on individual flows are set out in Table 5. The coefficients (Lagrange multipliers) for the Poland to Germany and Poland to UK flows are both equal to 0.42. This raises the adjustment factor for emigration from Poland from 10.64 (Table 4) to 18.31 (Table 5), while at the same time, the adjustment factor for Polish emigration to Germany and the UK falls to 7.69 (i.e., 18.31×0.42). For Polish immigration, the adjustment factor becomes smaller. The high adjustment factor for Polish receiving data was mainly a consequence of the big difference between the two figures for migration from Germany to Poland. Including a constraint for this flow raises the adjustment factor for Poland's reported flow from Germany by a factor of 1.74 (i.e., the adjustment factor of 14.25 is multiplied by 1.74 to get 24.80). In contrast, the adjustment factor for Poland's reported flow from the UK falls to 10.40 (i.e., 14.25×0.37). For the

Table 5 Estimates of adjustment factors for immigration and emigration, 2002–2007, including six additional constraints on individual flows

	Immigration	Emigration
Austria	1.17	1.35
Cyprus	0.88	4.71
Czech Republic	1.97	8.92
Germany	0.81	0.71
Denmark	0.72	0.74
Spain	0.73	4.32
Finland	1.18	1.12
Iceland	0.59	0.69
Italy	1.48	2.44
Lithuania	2.16	2.15
Luxembourg	5.45	2.08
Latvia	2.78	5.44
Netherlands	1.04	1.06
Norway	0.81	1.10
Poland	14.25	18.31
Sweden	1.00	1.10
Slovenia	4.90	2.33
Slovakia	8.34	39.40
United Kingdom	1.09	0.91
Coefficients for additional constraints (Lagrange multipliers)		
Immigration to Poland from Germany	1.74	
Immigration to Poland from the UK	0.37	
Emigration from Poland to Germany		0.42
Emigration from Poland to the UK		0.42
Emigration from Germany to the UK		1.70
Emigration from the Czech Republic to Slovakia		0.10

Czech Republic, the reported emigration numbers are considerably lower than the corresponding reported immigration numbers with one big exception: the number of emigrants reported to Slovakia is relatively large. Clearly, the emigration flows from the Czech Republic to all other countries need to be adjusted by a different factor than the emigration flow to Slovakia.

The adjustment factors in Table 5 illustrate how substantial improvements in the estimated adjustment factors can be made by introducing constraints on specific ‘problem’ flows in the matrix. For example, the inclusion of a constraint for the migration flow from the Czech Republic to Slovakia lowered the adjustment factor for Slovakia’s receiving migration data from 18.90 to 8.34. Another example is German’s receiving data. Here, the adjustment factor is reduced from 1.03 to 0.81. This is mainly explained by the reduction of the estimate of Polish emigration to Germany. Since Germany has a wide definition of migration, one would expect the adjustment factor to be below one. Thus, the adjustment factors in Table 5 appear more plausible than those set out in Table 4.

The harmonised migration tables that used the additional constraints are set out in Tables 6 and 7. The introduction of these constraints led to a further strong reduction in the differences between both tables, as indicated by the RMSE, which fell from 2131 to 952 or by a further 54%. To obtain a final single set of harmonised flows, we believe it is better to rely on Table 6 than on Table 7. This table gives more weight to the receiving country data, which we consider more reliable. Poulain, on the other hand, advocated taking the average of the two estimated matrices. This approach implies that the origin–destination patterns in the reported sending country data are as reliable as those in the reported receiving country data.

The average adjustment factors estimated for the period 2002–2007 (Table 5) can be applied to the annual reported migration data to create a time series of harmonised flows. In Fig. 1, the estimated total immigration and emigration flows for Germany from and to the other 18 countries in this study are compared. As expected, the estimated numbers are lower than the reported numbers because the definition for Germany is much wider than the harmonised definition. The figure also shows that estimated emigration increases more gradually over time than the reported numbers. In Fig. 2, the immigration and emigration flows for the UK are presented. Here, the average levels of the reported and estimated numbers do not differ much, but the estimated flows show a more gradual pattern over time than the reported flows. One reason for the sharp fluctuations in the reported numbers is that they are based on sample surveys.

6 Discussion

The aim of this paper has been to obtain a reasonable and consistent set of international migration statistics. For this purpose we have developed a model using statistical information from different countries. The method is based on an idea originally proposed by Poulain (1993, 1995). Our method differs from his in three important ways. First, we have estimated a set of adjustment factors for receiving and sending country data in a way that ensures consistency in the two sets of

Table 6 Estimated migration by country of origin and destination, including constraints on six individual flows, 2002/2007, based on numbers reported by receiving countries

From	To									
	AT	CY	CZ	DE	DK	ES	FI	IS	IT	LT
AT		36	610	11,526	217	563	129	20	1,144	37
CY	26		25	223	16	18	28	1	44	7
CZ	1,547	103		7,453	187	606	66	25	992	51
DE	18,148	291	2,679		2,864	11,631	1,092	152	18,920	1,057
DK	238	21	90	2,172		701	432	840	391	183
ES	822	39	139	11,887	1,259		763	40	3,020	544
FI	317	19	75	1,757	296	614		27	347	92
IS	37	0	8	191	1,192	95	59		51	22
IT	1,889	43	500	17,945	706	6,782	297	44		177
LT	211	31	92	3,635	741	1,655	86	162	558	
LU	79	2	5	1,845	116	89	59	16	315	11
LV	97	91	26	1,742	327	218	103	55	270	378
NL	930	61	502	11,060	619	3,465	309	33	1,336	89
NO	115	12	47	1,114	2,254	1,234	1,001	216	246	187
PL	6,146	659	3,163	110,701	1,744	6,024	221	1,324	13,361	259
SE	574	77	132	2,706	2,372	1,328	4,150	292	560	197
SI	653	8	32	1,454	33	99	7	5	473	4
SK	3,750	379	27,658	9,012	107	574	26	26	1,020	9
UK	1,436	2,780	996	10,722	2,493	28,145	1,121	135	6,725	1,886
Total	37,015	4,654	36,780	207,145	17,542	63,841	9,950	3,411	49,772	5,190

From	To									
	LU	LV	NL	NO	PL	SE	SI	SK	UK	Total
AT	44	26	584	90	2,570	307	488	1,730	1,522	21,642
CY	1	4	53	12	100	61	8	18	2,764	3,408
CZ	23	43	534	94	644	164	28	8,158	4,482	25,200
DE	2,475	462	9,586	1,843	71,514	3,374	1,462	3,715	20,770	172,034
DK	60	128	496	2,391	480	5,264	15	176	2,044	16,124
ES	132	49	3,237	624	1,698	1,300	41	297	15,907	41,798
FI	18	119	395	649	88	3,204	4	51	746	8,819
IS	2	17	78	303	162	462	6	18	455	3,157
IT	365	93	1 891	200	4,402	599	386	906	6,359	43,582
LT	7	656	316	752	610	574	1	39	2,734	12,858
LU		4	168	15	69	90	24	10	744	3,661
LV	9		131	189	90	264	0	38	1,338	5,367
NL	146	55		577	2,323	979	59	343	7,417	30,305
NO	12	68	473		679	5,098	2	201	1,818	14,778
PL	101	125	5,997	3,739		3,718	12	2,298	40,102	199,695
SE	74	149	727	3,996	1,608		74	168	3,505	22,690
SI	6	4	94	12	24	42		129	0	3,079
SK	24	12	486	193	254	110	20		5,001	48,662
UK	213	528	6,077	1,320	5,907	3,114	105	968		74,670
Total	3,715	2,540	31,321	17,000	93,222	28,723	2,735	19,264	117,708	751,530

Table 7 Estimated migration by country of origin and destination, including constraints on six individual flows, 2002/2007, based on numbers reported by sending countries

From	To									
	AT	CY	CZ	DE	DK	ES	FI	IS	IT	LT
AT		24	1,261	8,973	223	578	311	36	1,376	150
CY	26		99	266	26	87	56	0	181	42
CZ	1,658	116		4,999	213	309	253	21	998	89
DE	12,616	192	5,748		2,195	11,921	1,681	203	22,154	1,741
DK	168	17	132	1,924		1,229	271	991	527	482
ES	669	37	247	11,592	678		475	37	5,020	516
FI	109	25	47	852	449	754		60	228	24
IS	9	1	11	142	1,243	41	33		73	44
IT	1,433	15	164	24,877	364	3,675	331	41		26
LT	102	16	115	2,730	339	1,352	187	50	438	
LU	65	5	26	1,895	205	165	73	40	364	8
LV	98	44	30	1,640	245	100	250	27	277	748
NL	656	53	317	11,165	567	4,016	342	57	1,360	57
NO	76	18	47	780	3,399	867	940	453	160	118
PL	9,857	272	1,160	110,701	2,023	6,244	369	845	9,247	116
SE	327	80	114	1,789	3,460	1,477	3,728	453	507	52
SI	725	7	33	1,374	12	63	10	2	435	2
SK	6,974	46	24,784	10,028	144	617	20	0	1,642	0
UK	1,446	3,685	2,444	11,418	1,753	30,345	619	93	4,784	975
Total	37,015	4,654	36,780	207,145	17,542	63,841	9,950	3,411	49,772	5,190

From	To									
	LU	LV	NL	NO	PL	SE	SI	SK	UK	Total
AT	61	57	573	118	3,232	523	541	2,393	1,213	21,642
CY	10	86	45	11	520	59	0	148	1,745	3,408
CZ	25	58	721	140	5,203	213	70	8,158	1,956	25,200
DE	1,195	1,060	6,591	1,505	71,514	2,818	1,421	6,707	20,770	172,034
DK	101	233	443	2,170	613	3,868	22	70	2,863	16,124
ES	377	83	3,749	686	1,718	875	42	195	14,802	41,798
FI	80	30	262	873	70	3,617	5	11	1,321	8,819
IS	25	20	34	333	602	330	17	39	160	3,157
IT	531	20	1,294	295	1,015	484	368	96	8,551	43,582
LT	39	351	250	428	263	502	7	11	5,676	12,858
LU		9	201	26	47	152	10	22	345	3,661
LV	33		109	182	141	362	4	11	1,068	5,367
NL	203	35		777	1,085	958	48	146	8,462	30,305
NO	26	75	316		309	5,587	5	67	1,534	14,778
PL	412	61	10,205	2,319		5,539	34	189	40,102	199,695
SE	139	68	571	5,199	388		29	31	4,277	22,690
SI	57	1	70	11	11	88		13	164	3,079
SK	72	0	493	118	584	328	112		2,699	48,662
UK	328	294	5,395	1,809	5,907	2,420	0	956		74,670
Total	3,715	2,540	31,321	17,000	93,222	28,723	2,735	19,264	117,708	751,530

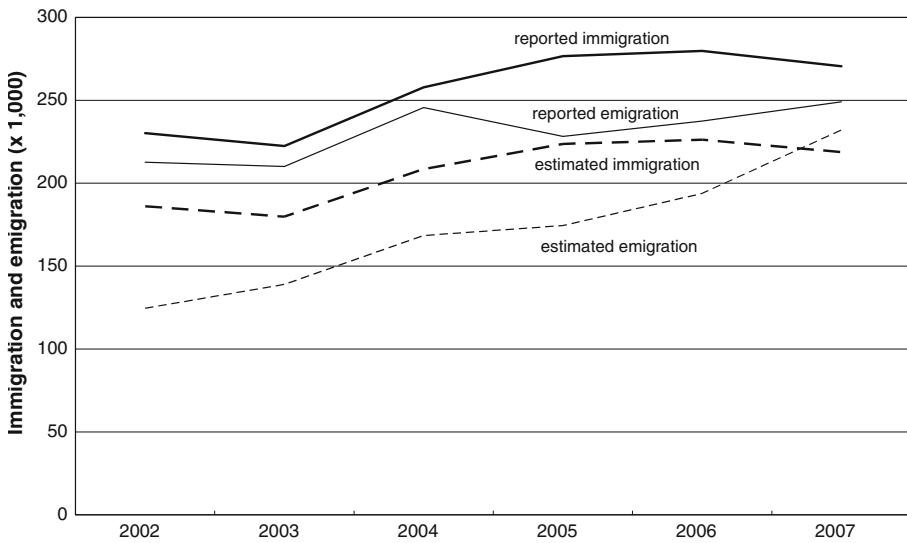


Fig. 1 Reported and estimated immigration from and emigration to 18 European countries, Germany, 2002–2007

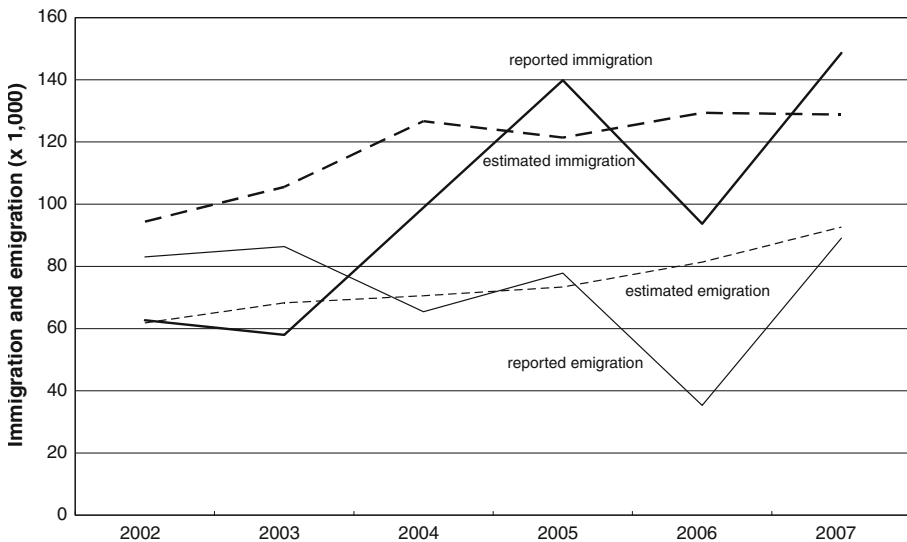


Fig. 2 Reported and estimated immigration from and emigration to 18 European countries, United Kingdom, 2002–2007

marginal totals. Second, we have introduced additional constraints on special origin–destination cases where the average adjustment factors do not apply. This allows us to include countries with less reliable data in our analysis. Third, instead of calculating the arithmetic average of the two estimated matrices, we believe it is

better to use the matrix giving more weight to the reported immigration numbers (i.e. Table 6). In this way we take advantage of the fact that the information on countries of origin in receiving country data tend to be more reliable than the country of destination information in sending country data. Finally, our estimates are consistent with the harmonised migration definition based on an (intended) minimum duration of stay of 12 months.

Due to differences in definition, coverage and registration, the origin–destination matrix of migration flows between European countries based on receiving country data tends to differ from the matrix based on sending country data. Germany has a wide definition of migration, as it does not include a time constraint and thus the reported number may well include short term migrants. In contrast, Poland has a very narrow definition of migration and, as a consequence, the reported numbers are very low. By comparing corresponding reported immigration and emigration flows for 19 European countries, we have assessed to what extent German migration statistics are higher than they would be under a harmonised definition and to what extent Polish migration statistics are lower.

However, the large differences between European countries cannot be explained by differences in definitions alone. First, these differences cannot explain why emigration flows are more likely to be underestimated than immigration flows. Second, whereas 11 countries employ a duration limit that is shorter than that of the harmonised definition (Kupiszewska and Wisniowski 2009), only five of these countries have an adjustment factor of immigration below one. The other six countries with durations of 6 months or shorter have adjustment factors for immigration greater than one. These include Austria, Czech Republic, Italy, Luxembourg, the Netherlands and Slovenia. Thus, to an important extent, the differences must also be caused by problems of coverage. This is confirmed by a study comparing migration statistics between Sweden, Denmark and Belgium which suggests that less than 25% of differences are due to differences in the duration criterion (Nowok et al. 2006). The effects of differences in definition and coverage may offset each other to some extent. One would expect the under-registration of short term migrants to exceed that of long-term migrants. A wide definition of migration (i.e. a short duration of stay) would lead to a higher reported number of migrants than would be expected on the basis of the harmonised definition. Under-registration, however, would lead to a smaller number. This may explain why the adjustment factors for Germany are not as low as one might expect from applying the very wide definition.

The main reason for the relatively low numbers reported by sending countries is that emigrants do not have strong incentives to report leaving a country. In particular, this applies to EU citizens who can live in another EU country without asking for a residence permit. One solution might be to introduce a removal card system (Nowok et al. 2006). Here, any person leaving country *A* would be required to fill in a form to be given to the authorities in country *B* at arrival. After country *B* has determined whether or not the person is an international migrant under a harmonised definition, it would then inform country *A* of the arrival. The Nordic countries have such a system and their immigration and emigration statistics are mutually consistent (Herm 2006a). However, policy makers tend to be more

interested in migrants from outside Europe and asylum seekers than intra-European migrants, and therefore such a system is not likely to have a high priority in the future. As long as such a system is lacking, cross-country comparability of migration statistics can only be achieved by comparing statistics from different countries. To the extent that the differences between countries are caused by differences in definitions and coverage, the differences may be expected to remain systematic over time. The method developed in this paper aims to assess the size of these systematic differences. Table 5 shows that for 10 out of the 19 countries in this study, the adjustment factor for sending country data exceeds two, meaning that reported emigration numbers are underestimated by more than 50% in relation to the 1-year duration definition. As a consequence, reported net migration totals may be overstated.

In addition to ‘correcting’ the reported receiving and sending country migration data for differences in definition and coverage, our method contributes to producing estimates that tend to fluctuate less strongly over time. One clear example concerns the UK. Since the UK uses a general purpose passenger survey, the reported flows fluctuate considerably over time. Moreover, flows to some (smaller) countries may not be observed in some years. We believe our method produces more stable estimates of migration flows for the UK (and other countries relying on sample data). Interestingly, the estimated adjustment factors for the UK are close to one. This implies that the sample survey used for estimating migration to and from the UK provides a reasonably reliable estimate of total migration flows on average, but that the annual estimates are affected by sizeable random fluctuations.

The adjustment factors shown in Table 5 can be used to adjust migration numbers to and from countries not included in the matrix, so that total immigration and emigration numbers and total net migration can be estimated for the 19 countries in this study. Before doing so, one first has to make sure that the share of unknowns in the migration statistics can be distributed evenly across all origins or destinations. If so, the adjustment factors will take this into account. Thus, for estimating total immigration and emigration numbers, the adjustment factors should be applied to total migration numbers excluding unknowns.

The matrix may be extended to include flows with missing data. Raymer (2008) developed a two-step estimation method for countries with missing data (see also De Beer et al. 2009; Raymer and Abel 2008). The first step estimates missing immigration and emigration totals based on harmonised migration flows and covariate information. The second step uses the origin–destination interaction patterns of the harmonised migration flows and covariate information to estimate the missing interaction patterns. This estimation step takes into account the fact that migration is relatively high, for example, between neighbouring countries and countries belonging to a similar language group.

Finally, work is currently being carried out to integrate harmonisation and estimation of missing data into a single (Bayesian) model that also includes measures of uncertainty and expert judgements. The Integrated Modelling of European Migration (IMEM) project recently funded by New Opportunities for Research Funding Agency Co-operation in Europe (NORFACE) is expected to develop such a model (see <http://www.norface.org/migration12.html>) over the next

couple of years. We hope this study will provide an important foundation for work such as this, and other projects aiming to improve our knowledge and understanding of the complexity of international migration.

Acknowledgments This work was supported by a European Commission/Eurostat tender no 2006/S100 106607 (OJ 27/05/2006) for the supply of statistical services. The authors would like to thank the other team members of the MIMOSA (MIgration MOdelling for Statistical Analyses) project for their comments and suggestions on earlier efforts (see <http://mimosa.gedap.be>).

Open Access This article is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Abel, G. J. (2009). *International migration flow table estimation*. PhD thesis, University of Southampton, School of Social Sciences.
- Bilsborrow, R. E., Hugo, G., Oberai, A. S., & Zlotnik, H. (1997). *International migration statistics: Guidelines for improving data collection systems*. Geneva: International Labour Office.
- Champion, A. G. (1994). International migration and demographic change in the developed world. *Urban Studies*, 31(4/5), 653–677.
- De Beer, J., Van der Erf, R., & Raymer, J. (2009). Estimates of OD matrix by broad group of citizenship, sex and age, 2002–2007 Report for MIMOSA-project. Available at http://mimosa.gedap.be/Documents/Mimosa_2009b.pdf.
- Fassmann, H. (2009). European migration: Historical overview and statistical problems. In H. Fassmann, U. Reeger, & W. Sievers (Eds.), *Statistics and reality. Concepts and measurements of migration in Europe* (pp. 21–44). Amsterdam: Amsterdam University Press.
- Herm, A. (2006a). Recommendations on international migration statistics and development of data collection at an international level. In M. Poulain (Ed.), *THESIM: Towards harmonised European statistics on international migration* (pp. 77–106). Louvain-la-Neuve: Presses universitaires de Louvain.
- Herm, A. (2006b). Country report Sweden. In M. Poulain (Ed.), *THESIM: Towards harmonised European statistics on international migration* (pp. 633–643). Louvain-la-Neuve: Presses universitaires de Louvain.
- Kelly, J. (1987). Improving the comparability of international migration statistics: Contribution by the Conference of European Statisticians from 1971 to date. *International Migration Review*, 21(4), 1017–1037.
- Kraier, A., Jandl, M., & Hofmann, M. (2006). The evolution of EU migration policy and implications for data collection. In M. Poulain (Ed.), *THESIM: Towards harmonised European statistics on international migration* (pp. 35–75). Louvain-la-Neuve: Presses universitaires de Louvain.
- Kraly, E. P., & Gnanasekaran, K. S. (1987). Efforts to improve international migration statistics: A historical perspective. *International Migration (Special issue: Measuring international migration: Theory and practice)*, 21(4), 967–995.
- Kupiszewska, D., & Nowok, B. (2008). Comparability of statistics on international migration flows in the European Union. In J. Raymer & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates* (pp. 41–71). Chichester, England: Wiley.
- Kupiszewska, D. & Wisniowski, A. (2009). Availability of statistical data on migration and migrant population and potential supplementary sources for data estimation Report for MIMOSA project. Available at http://mimosa.gedap.be/Documents/Mimosa_2009.pdf.
- Nowok, B., Kupiszewska, D., & Poulain, M. (2006). Statistics on international migration flows. In M. Poulain (Ed.), *THESIM: Towards harmonised European statistics on international migration* (pp. 203–232). Louvain-la-Neuve: Presses universitaires de Louvain.

- Official Journal of the European Union (2007) Regulation (EC) No 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers. Available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:199:0023:0029:EN:PDF>.
- Poulain, M. (1993). Confrontation des Statistiques de migrations intra-européennes: Vers plus d'harmonisation? *European Journal of Population*, 9(4), 353–381.
- Poulain, M. (1995). Towards a harmonisation of migration statistics within the European Community. In S. Voets, J. Schoorl, & B. De Bruijn (Eds.), *Demographic consequences of international migration* (pp. 11–25). The Hague: Netherlands Interdisciplinary Demographic Institute.
- Poulain, M. (1999). International migration within Europe: Towards more complete and reliable data? Working Paper 12, joint ECE-Eurostat Work Session on Migration Statistics, Geneva, Switzerland.
- Poulain, M. & Dal, L. (2008). Estimation of flows within the intra-EU migration matrix. Report for the MIMOSA project Available at http://mimosa.gedap.be/Documents/Poulain_2008.pdf.
- Poulain, M., Perrin, N., & Singleton, A. (Eds.). (2006). *THESIM: Towards harmonised European statistics on international migration*. Louvain-la-Neuve: Presses universitaires de Louvain.
- Raymer, J. (2008). Obtaining an overall picture of population movement in the European Union. In J. Raymer & F. Willekens (Eds.), *International migration in Europe: Data, models and estimates* (pp. 209–234). Chichester, England: Wiley.
- Raymer, J. & Abel, G. (2008). The MIMOSA model for estimating international migration flows in the European Union. Joint UNECE/Eurostat Work Session on Migration Statistics, Working paper 8. Geneva: UNECE/Eurostat. Available at: <http://www.unece.org/stats/documents/ece/ces/ge.10/2008/wp.8.e.pdf>.
- Raymer, J., & Willekens, F. (Eds.). (2008). *International migration in Europe: Data, models and estimates*. Chichester, England: Wiley.
- Thierry, X., Herm, A., Kupiszewska, D., Nowok, B., & Poulain, M. (2005). How the UN recommendations and the forthcoming EU regulation on international migration statistics are fulfilled in the 25 EU countries? Paper presented at the XXV International Population Conference, Tours, 18–23 July 2005. Available at <http://iussp2005.princeton.edu/download.aspx?submissionId=51414>.
- UNECE. (2009). Improving migration statistics by exchange of data between countries. 95th DGINS conference “Migration—Statistical mainstreaming”, 1 October 2009, Malta. Available at http://epp.eurostat.ec.europa.eu/portal/page/portal/conferences/documents/95th_dgins_conference/UNECE.pdf.
- United Nations. (1998). Recommendations on statistics of international migration. Statistical Papers Series M, No. 58, Rev.1, Department of Economic and Social Affairs, Statistics Division, United Nations, New York. Available at http://unstats.un.org/unsd/publication/SeriesM/SeriesM_58rev1E.pdf.
- United Nations. (2002). *Measuring international migration: Many questions, few answers*. United Nations, New York: Population Division. Department of Economic and Social Affairs.
- Van der Erf, R. & Van der Gaag, N. (2007). An iterative procedure to revise available data in the double entry migration matrix for 2002, 2003 and 2004. Discussion Paper, Netherlands Interdisciplinary Demographic Institute, The Hague. Available at http://mimosa.gedap.be/Documents/Erf_2007.pdf.
- Willekens, F. (1994). Monitoring international migration flows in Europe: Towards a statistical data base combining data from different sources. *European Journal of Population*, 10, 1–42.
- Willekens, F. (1999). Modeling approaches to the indirect estimation of migration flows: From entropy to EM. *Mathematical Population Studies*, 7(3), 239–276.
- Zlotnik, H. (1987). The concept of international migration as reflected in data collection systems. *International Migration Review*, 21(4), 925–946.